DOCUMENT RESUME

ED 398 280 TM 025 483

AUTHOR Fenster, Mark J.
TITLE An Assessment of "Middle" Stakes Educational Accountability: The Case of Kentucky.
PUB DATE Apr 96
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Accountability; *Achievement Gains; Educational Assessment; Educational Change; Elementary Secondary Education; Equal Education; *Incentives; Political Influences; State Legislation; State Programs; *Test Construction; *Testing Problems; Testing Programs; Test Use
IDENTIFIERS *Kentucky Education Reform Act 1990; Kentucky Instructional Results Information System; *Reform Efforts

ABSTRACT

In 1990 the Kentucky state legislature passed the Kentucky Education Reform Act (KERA), which mandated a total overhaul of the state's kindergarten through grade 12 public school system and was designed to result in equitable education for all students. Accountability components of the KERA include financial incentives for staff in schools where student gains are exemplary and the use of sanctions to cause staff in ineffective schools to boost student achievement gains to an acceptable level. These features are operationalized through the Kentucky Instructional Results Information System (KIRIS). KIRIS had to comply with the legislative mandates and had to produce assessments that would be technically defensible and politically credible for "middle" stakes, rather than high- or low-stakes assessment. The state's reform efforts provide an opportunity to examine systemic reform in assessment and accountability. This description of assessment development and implementation demonstrates that the KERA has had classroom impact. Change has been forced in many areas because the incentive and sanction provisions have made KERA and KIRIS impossible to ignore. Now the state's problem is to move on to a reform effort that can fine tune itself without the external shocks that came with KERA implementation. In the current political climate, the system sustained on rewards and sanctions may not last until its supposed end date in 2012. (Contains 2 figures, 3 tables, and 20 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 398 280

AN ASSESSMENT OF "MIDDLE" STAKES EDUCATIONAL ACCOUNTABILITY: THE CASE OF KENTUCKY

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
MARK FENSTER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

by

Mark J. Fenster
Wedgewood Associates
931 Austin St. Apt. 2
Kalamazoo, MI 49008-1104 USA

mfenster@edcen.ehhs.cmich.edu

Paper presented at the Annual meeting of the 1996 American Educational Research Association, New York City, April 12, 1996.

BEST COPY AVAILABLE

TM025483

Background Information

This paper should be considered in the environment of educational reforms under way in Kentucky since 1990. The scale of these reforms is massive and unprecedented for any state. In 1989 the Kentucky Supreme Court declared the Commonwealth's existing rules and procedures for financing schools and delivering educational services to be unconstitutional. In 1990 the state legislature passed the Kentucky Education Reform Act (KERA), which mandated a total overhaul of the K-12 public education system and was designed to result in equitable educational services to all students.

The main features of the KERA are (1) prescribed statewide academic expectations; (2) use of a model curriculum framework; (3) a commitment to helping all children to become proficient in performing rigorous state standards that emphasize application of what is learned; (4) heavier concentration of learning resources on students who are not learning up to their potential; (5) extensive parental involvement; (6) site-based management of schools; (7) use of financial incentives to reward staff in schools where student gains are exemplary; and (8) use of sanctions, including the assistance of distinguished educators, to cause staffs in ineffective schools to bring student achievement gains to an acceptable level. These last two features of KERA were operationalized through an integrated, comprehensive assessment system, The Kentucky Instructional Results Information System (KIRIS).

KIRIS *had* to comply with legislative mandates, which are evolving; *had* to provide performance measures; and *had* to produce assessments that would be technically defensible and politically credible for making "middle" stakes decisions on rewards and sanctions to schools. The Kentucky Department of Education *had* to develop KIRIS a couple of years *before* the field of educational measurement updated the standards for judging assessment systems (Linn, 1994). The educational measurement profession is in the process of updating its standards. The current Standards for Educational and Psychological Tests (APA, 1985) do not deal with the "middle stakes", school based, Kentucky assessment system very well.

Beyond a single state case study, systemic reform initiatives have become more common in recent years. Kentucky has gone further in systemic reform than most other states in the United States. For this reason, Kentucky is seen as a bellwether for the country to examine the practicality of systemic reform initiatives for K-12 public education. Educational researchers can learn from Kentucky's experience with systemic reform, let alone Kentucky's unique approach to assessment and accountability.

Brief History of Education in Kentucky Before and After the Passage of KERA

The Commonwealth of Kentucky used the Kentucky Essential Skills Test (KEST) in the middle 1980s and the Comprehensive Test of Basic Skills (CTBS-IV) in 1988-1989 and 1989-1990 to assess students. The Commonwealth could take over school districts if their students did not perform satisfactorily on KEST. However, some people in the state thought there were some fundamental problems with an accountability system focused at the district level. Within a district, schools with weak or descending test scores could be counterbalanced by other schools with strong and improving test scores in that district. This problem led state leaders to reconceptualize accountability so that it applies to the school, rather than at the district level.

In June of 1989, the Kentucky Supreme Court ruled the public school system in the Commonwealth was unconstitutional. Based on the evidence presented in *Rose v. the Council for Better Education*, (1989), the court concluded that each child in the Commonwealth was **not** being provided with an equal opportunity to have an adequate education. The inequities between rich and poor school districts were too large, depriving children in poorer districts a fair and equal

opportunity to receive an adequate education. According to the court, the responsibility for providing an adequate education for **all** children of the Commonwealth rested with the General Assembly. In response to the court order, the state legislature passed the Kentucky Education Reform Act of 1990 (or KERA).

KERA includes a number of legislative mandates, two of which are described here. One mandate is that a primarily performance-based assessment procedure be used. Instead of using only multiple-choice questions as did KEST and CTBS-IV, KERA required the Kentucky Department of Education (KDE) to assess what "students could do with what they know." As a result of this mandate, KDE designed and has been developing the performance assessment component of KERA. This assessment system is named the Kentucky Instructional Results Information System (KIRIS).

KERA also mandated that the assessment system (KIRIS) must be usable for granting rewards to schools that have an increased proportion of successful students and for delivering sanctions to schools that have a decreased proportion of successful students. As documented in Guskey (1994, p.81),

> The legislation requires the State Board to establish . . . a threshold level for school improvement . . . to determine the amount of success needed for a school to receive a reward. The threshold definition shall establish the percentage of increase required in a school's percentage of successful students, as compared to a school's present proportion of successful students, with consideration given to the fact that a school closest to having one hundred percent (100%) successful students will have a lower percentage increase required.

KERA further requires that school success shall be determined by measuring a school's improvement over a two year period. As discussed in Guskey (1994, p. 82), a school that does not reach its prescribed threshold level

> . . . but maintains the previous proportion of successful students shall be required to develop a school improvement plan and shall be eligible to receive funds from the school improvement funds pursuant to KRS 158.805. A school in which the proportion of successful students declines by less than five percent (5%) shall be required to develop a school improvement plan, shall be eligible to receive funds from the school improvement fund, and shall have one or more Kentucky distinguished educators assigned to the school to carry out the duties as described in KRS 158.782. A school in which the proportion of successful students declines by five percent (5%) or more shall be declared by the State Board for Elementary and Secondary Education to be a 'school in crisis,' and the State Board is to implement defined sanctions.

The rewards and sanctions make the KERA reform a "stakes" program. If the percentage of successful students increase, employees in the school (the principal, teachers, support staff, and others) were eligible for financial awards (about $1200 in the first biennium). If the percentage of successful students decreased, schools were given additional assistance in the form of distinguished educators. In the early years of the Kentucky assessment program, there were no strongly negative sanctions associated with poor performance on KIRIS. We note here that the most severe sanction, the school in crisis sanction, has not yet been implemented. Without the school in crisis sanction, we find it hard to classify the KIRIS assessment system as "high stakes". For this reason, we classify the Kentucky assessment system as a "middle stakes" program.

Description of the KIRIS Assessment

In order to understand some of the issues discussed in this paper, it is necessary to have a rudimentary understanding of the elements of the KIRIS assessment program. The purpose of this section is to provide the reader with this rudimentary information.

The accountability index. The basis for describing a school's accomplishment is the KIRIS accountability index. The accountability index for a school is the average performance of the school's students over six separate measures: five cognitive achievement measures and one noncognitive achievement measure. Each cognitive achievement measure reflects a school's performance in one curriculum area. A school's performance on each of the six measures is also reported. In a rough (but imprecise) way, a school's score on the accountability index can translate into percentage of successful students specified by KERA.

Cognitive achievement at the school level. The measure of a school's accomplishment in each cognitive achievement area is the average achievement score of its students. For each of the five curriculum areas, a student's score is obtained as follows. As a result of several types of assessments in an area (which are described later), each student is classified into one of four quality levels: novice, apprentice, proficient, and distinguished. Next each student is assigned points on the KIRIS score scale as shown in Table 1.

Table 1: Relationship Between Level of Student Performance and Accountability Score Scale Points

Student Quality Level	Corresponding Accountability Score Scale Points
Novice	0
Apprentice	40
Proficient	100
Distinguished	140

A student's score on the cognitive dimensions has a possible range from 0 to 140. If a student is absent from the assessment, the student is assigned a KIRIS score scale of 0.

After assigning points for a curriculum area to each student, all of the students' scores are averaged. The average is calculated separately for grades 4, 8, and 11 (previously 12). The process is repeated for each of the five curriculum areas. These averages are a school's cognitive measures.

A school's noncognitive achievement. In addition to the five cognitive measures, each school receives a noncognitive measure. This measure is a composite of a school's attendance, retention, dropout, and transition to adult life assessments (dropout applies only to middle schools and high schools and transition to adult life applies only to high schools.) The highest possible score a school may attain on the noncognitive assessment is 100.

A school's accountability index. The single number by which a school is judged is the accountability index. This index is the school's average performance over all the cognitive and noncognitive measures. This sum is then divided by six.

Although the theoretical range of the index that combines cognitive and noncognitive assessments is 0 to 133.3, the extremes of 0 and 133.3 cannot be attained except in extraordinary circumstances. For example, in order for a school's accountability index to equal 0, all students in

BEST COPY AVAILABLE

a school must score at the novice level and the school must have a score of 0 on the noncognitive measure. Similarly, at the other extreme all students must score at the distinguished level and the school's noncognitive measure must equal 100. (Interested readers may find a more complete description of the computation of the accountability index for the first biennium (1992-1994) in Guskey [1994].)

The desired minimum score for a school. Achieving at least the proficient level for all students is a school's goal. This qualitative goal can be translated to a quantitative accountability index value. Since the proficient level translates to 100 on the KIRIS score scale, an average student score of 100 in each cognitive area is the desired minimum accountability score. One of the unstated goals of the Kentucky educational reform movement is that each Kentucky school is supposed to reach the desired minimum score of 100 in 20 or fewer years.

How a school receives rewards or gets sanctioned. Based on the 1991-92 KIRIS assessment, schools received a baseline score on the accountability index. The baseline score was subtracted from 100 (the desired minimum accountability score 20 years after the start of the program). The difference between the desired accountability score (100) and the baseline score was a gap that each school had to close. The gap between the desired accountability score and the baseline score was divided by 10. The division by 10 represents the length of the program--10 bienniums, or 20 years. The gap divided by 10 represented the average gain on the accountability index needed by a school to avoid sanctions.

An example may clarify the previous paragraph. Assume a school received an accountability score of 40 on the baseline (1991-92) KIRIS assessment. This baseline score (40) would be subtracted from 100. The difference between the desired accountability score (100) and the baseline score (40) is a 60 point gap. This gap would be divided by 10, to account for the length of the program. In this case, the gain on the accountability index needed by this school to avoid sanctions is 6 points. The school's 1992-93 and 1993-94 KIRIS accountability index results would be averaged to determine the school's accountability index value at the end of the first biennium.

If the school in this example had a biennium accountability average of 46, the school would neither be sanctioned or rewarded. If the school's accountability index average was 47 or higher, the school would receive financial rewards. If a school's accountability index average was less than 40, the school would face sanctions under the KERA legislation. If the school's average accountability index was between 46 and 47, the school would be classified as successful. Such a result would subject the school to neither rewards or sanctions. The KERA legislation did not clearly define what happens to a school such as this one if its average accountability index value was between its baseline and its threshold (in this example, between 40 and 46). The Kentucky Department of Education (KDE) has determined that if a school does not achieve its threshold (46 in this case), but increases its accountability score, that school needs to develop a school improvement plan.

How the Assessment Tasks Within a Cognitive Measure are Weighted

Since the initial year (1991-1992) cognitive measures were obtained in the curriculum areas of mathematics, reading, science, social studies, and writing. Each area is assessed by a variety of formats that are weighted differently. In 1993-1994, the formats for the first biennium were weighted as shown in Table 2.

BEST COPY AVAILABLE

Table 2: Cognitive Weights on the KIRIS Assessment 1992-1994

	Component		
Assessment Format	Social Studies Science Math	Reading	Writing[1]
1. Open-Ended Common Questions (Five Questions)	40%	50%	NA
2. Open-Ended Matrix-Sampled (Each student is randomly assigned to answer 2 of a pool of 24 questions.)	40%	50%	NA
3. Performance Events	20%	NA	NA
4. Multiple Choice Questions[2]	0%	0%	NA
5. Portfolio[3]	0%	0%	100%
6. On-Demand Writing Prompts[4]	NA	NA	0%
TOTAL	100%	100%	100%

How the Components Within the Noncognitive Measures are Weighted

The weights of the components comprising the noncognitive measures are shown in Table 3.

Table 3: Noncognitive Weights on the KIRIS Assessment, 1992-1994

Component	4th grade	8th grade	High School
Attendance	80%	40%	20%
Retention	20%	40%	5%
Dropout	NA	20%	37.5%
Transition to Adult Life	NA	NA	37.5%
TOTAL	100%	100%	100%

We note here that starting with the 1994-1995 KIRIS assessment, the noncognitive measure will be lagged by one year. That is, the computation for the 1994-1995 school year will be based on a school's 1993-1994 data for the four components. This was done because there was insufficient time to collect and disseminate the data for the 1994-1995 school year.

---

1. Writing is assessed only through a portfolio.
2. Multiple-choice tests were administered, but the results were not counted in the accountability measure.
3. Only mathematics was assessed by a portfolio, and the results were not counted toward the accountability measure.
4. On-demand writing prompts were administrated, but the results were not counted in the accountability measure.

Recent Changes in the Assessment

Because the KIRIS assessment is an innovative and developing program, it is reasonable to expect that changes and fine-tuning will be done each year the program is in place. The following changes were made to the KIRIS assessment during the 1994-1995 school year.

1. Assessment using a mathematics portfolio at the fourth grade was discontinued and in its place, a fifth grade mathematics portfolio was used.

2. Whereas the mathematics portfolio had not been counted in the accountability index in the past, it counted in the 1994-1995 assessment. The State Board for Elementary and Secondary Education decided that the mathematics portfolio counted for 30% of the total math score.

3. The accountability grade in high school was grade 11 instead of grade 12. However, the portfolios in writing and mathematics will still be due in grade 12.

4. The accountability grades were 4, 8, and 11, However, grade 12 students were also assessed. This was done for purposes of equating scores from previous years to the 1994-1995 year.

5. Although curriculum areas beyond the five mentioned previously were assessed in 1993-1994, they did not count in the accountability index. These areas were arts, humanities, and practical living/vocational studies. These assessments will count in the second biennium (1994-1996). Currently, arts, humanities, and practical living/vocational studies are assessed in the scrimmage (practice) tests that can be administered in non-accountability grades. Performance events were also used for these areas in 1993-1994.

In what may become the biggest change to the KIRIS assessment, a new RFP was issued by the KDE in late 1995. With this RFP, the KIRIS assessment will be redesigned, heading into the new millennium. Like the earlier KIRIS assessment, most cognitive areas will be covered with multiple modes of assessment; multiple choice, constructed response portfolio and performance events. However, multiple choice questions may count on the KIRIS assessment for the first time, increasing the number of assessment modes that carry weight to determine a school's accountability score. It is possible that the new KIRIS assessment will show improved reliability with both the individual student level and the school level scores.

Intent of KERA: Shake Up the System

KERA was a complex piece of legislation that accomplished more than simply mandating a performance assessment system. As envisioned by its supporters, KERA would create an environment for an inclusive system focusing on improved educational outcomes for school children. Borrowing from the African proverb that "it takes an entire village to raise a child", teachers, principals, and parents would all work **together** in the Site-Based Management Councils (SBMC) to improve instruction, help raise educators' expectations concerning the work students were capable of doing, and (eventually) move the school to higher KIRIS scores. Inclusive models of professional development, and simultaneously involving many different stakeholder groups in public education might help make a difference in student outcomes. Currently, much professional development activity in the United States focuses on one group at a time, asking "what do the parents have to do to improve student outcomes", "what do the teachers have to do to improve student outcomes", or "what do the principals have to do to improve student outcomes". These "one group at a time" efforts often work in isolation from the professional development of other major stakeholder groups in public education. With the SBMC, Kentucky's systemic reform

initiative has a mechanism for stakeholder groups to work together for improved education outcomes at the local level. Taken together, the eight propositions that constitute the philosophy of KERA (Wilkerson and Associates, Inc. 1994) were designed to shake up public education in Kentucky. The eight propositions are:

A. All children can learn at a relatively high level.
B. The state should set high standards of achievement for all children.
C. More learning resources should be focused on students who have not succeeded in meeting the state's learning standards.
D. Decisions affecting instruction can best be made at the local school level.
E. In the primary schools, students should not be labeled as belonging to a specific grade level.
F. It is not enough to require that students show their knowledge of facts; they must also demonstrate that they can apply what they know in real life situations.
G. Both rewards and sanctions are necessary to hold schools accountable for improving student performance.
H. Higher performance levels by all children are important for economic growth of Kentucky.

Five of the propositions (A, B, D, F and H) were taken from the systemic education reform literature (Cohen 1995). However, instead of talking about systemic educational reform, Kentucky overhauled its K-12 public education, installed new assessments, introduced an accountability system, and made a major financial commitment to professional development of educators. KERA was passed in response to a court order and a perceived crisis in the quality of public education in the Commonwealth. KIRIS was born out of frustration with an incremental approach to reform that never seemed to change the status quo (Haertel 1994). The intent of KERA was to shock the system, to make a radical break with educational business as usual, and to set a new course. As noted by Haertel (1994)

> The legislature soberly recognized that new tests alone would not be enough and embedded KIRIS in a comprehensive reform package. Educational funding was increased dramatically to pay for additional teacher training, new educational materials, and other needed changes. In addition, the legislature deserves credit for recognizing that change will take time. Many promising educational innovations have been undermined by an insistence on quick results. KIRIS features a reasonable phased implementation and a realistic 20-year time line for attaining its ultimate goals. Along the same lines, the Kentucky reforms take account of the different starting points of various schools, differentiating improvement targets according to initial achievement levels.

Impact of the KIRIS assessment on education in Kentucky

KERA called for extremely broad and sweeping educational reforms. It has been six years after the passage of KERA, and five years after the implementation of the KIRIS assessment. Proponents of KERA reasoned "assessment would drive curriculum". Educators (teachers, principals, superintendents) could not ignore the Kentucky reform effort because of the sanctions component of the KIRIS assessment. If educators ignored the reform effort by conducting business as usual, and KIRIS assessment scores went down, the schools where those educators worked would be subject to sanctions. If the school was subject to the most severe sanction, the school in crisis component, teachers and principals *might* lose their jobs. However, if educators ignored the reform effort by conducting business as usual and KIRIS assessment scores went up, the State would be satisfied. KERA mandated that assessment scores increase. Teachers, with the help of principals and site based management councils, were responsible for providing instruction

BEST COPY AVAILABLE

that would increase assessment scores. Neither KERA, nor the KDE has mandated a teaching style or a structured curriculum.

Proponents of KERA thought the rewards and sanctions component of the KIRIS assessment was required for educators to take to reform seriously. Additionally, proponents of KERA thought that the "stakes" associated with the assessment would play a crucial component in the systemic reform initiative, driving instruction and curriculum towards tasks stressed on KIRIS. With the reforms in place for six years, it may be useful to determine the impact of the KIRIS assessment in reforming education in Kentucky. Clearly, long term impacts of the assessment system may not be noticed for many years.

Positive Effects of the Assessment

1. Increased concentration on students' writing. As envisioned by KERA supporters, the KIRIS system of rewards and sanctions is supposed to motivate principals, teachers, and site-based management councils to alter the instructional curriculum presented to students. It has been argued in Kentucky that "assessment drives curriculum." The KIRIS assessment is heavily oriented toward writing. The biggest weight on the KIRIS assessment is assigned to the open-ended, constructed response questions. The second biggest weight is assigned to the writing portfolio. In the five cognitive areas covered in the first biennium, writing in one form or another accounted for 88 percent of the weight on the KIRIS assessment. Counting performance events as writing (students have to write their responses to situations) increases the weight of writing to 100 percent of the KIRIS assessment. Even in subjects like mathematics, students must write answers to questions. Students who do not write well will not do well on the KIRIS assessment, irrespective of their knowledge of subject matters like mathematics. Such a heavy weight on writing on the assessment may have made it easier for educators to stress writing when teaching. Students do report more writing under the reforms (Coe, Leopold, Simon, & Williams, 1994).

2. Improvement in students' writing quality. Scores on the writing portfolios have improved since the baseline KIRIS assessment in 1991-1992. Teachers, District Assessment Coordinators, and superintendents report almost unanimously that writing has improved; and the writing improvement was over and above what would have been expected of most school children of the same age. We *believe* (but we cannot be sure) that the reported increase and improvement in students' writing is due to the heavy weight on writing on the KIRIS assessment and, ultimately, the prospect of rewards and sanctions based on KIRIS assessment results. It is a limitation of this paper that we did not gather and study student portfolios and other evidence to assess whether the quality of students' writing has actually improved.

3. Involvement of students in cooperative problem-solving. Performance events add a group component to the KIRIS assessment. However, performance events accounted for 12 percent of the weight of the KIRIS assessment in the first biennium. Even with a relatively light weight, the performance events (group activities involving problem solving or experimentation) meaningfully engage students. Consistent with the 1994 revised legislative requirements, they yield experience but not assessments of the ability of individual students to work productively and collaboratively in groups and perform important learning targets. Students reported increased group work since the passage of the KERA (Coe, Leopold, Simon, & Williams, 1994).

4. Instructional contribution of portfolios. Another benefit of the KIRIS assessment has been the development of portfolios. Portfolios of students' mathematics and written work appear to have great instructional potential. Students have to choose their best work to put in the portfolio, and this gives students a chance to reflect upon their intellectual growth over a school year. A more passive assessment system, like a test consisting of 100 percent multiple-choice questions, does not meaningfully engage the student in the same way. Choosing pieces for inclusion in the portfolio engages the student more, thereby increasing the student's involvement in deciding the

material on which the student is to be assessed. For teachers, evaluating the best work of **their** students gives them critical feedback that they may use in deciding what instructional materials and assistance would best serve the entire class as well as individual students.

5. Increased Attention to Students with Special Needs. The limited evidence we gathered with respect to special populations showed that the educational system was paying considerably more attention to these groups than before the KIRIS system was set up. Special education students always take the KIRIS assessment unless they are severely handicapped. A severely handicapped student is assessed using an alternative portfolio. Some educators reported that the educational system was paying more attention to special education students because such students were included in the accountability system. Nearly 99.5% of students in the accountability grades are assessed in one form or another by KIRIS. Many (if not most) other statewide assessment programs do not include special education students, and other students with disabilities. It is possible that school systems would take the difficult job of educating children with disabilities more seriously if such students were included in an assessment program. If KDE can document that special education students are receiving increased instructional attention because of KIRIS, this evidence would support the consequential validity of the accountability index.

6. The Sanctions Component of KIRIS May Have Provided Members of the State Legislature With Some Protection Against Retaliation from Constituents after supporting a Big Tax Increase. There is some evidence from national public opinion polls that the public will tolerate large tax increases if the money raised from such a tax does not "go into a black hole" (Johnson and Immerwahr, 1994). In other words, if the public feels the additional money raised by a tax increase is not wasted, politicians supporting a tax increase are less likely to face retaliation at the voting booth in future elections. The accountability provisions of KIRIS may have provided members of the state legislature who voted for the tax increase to fund KERA a buffer against constituents generally opposed to tax increases. Members of the State Legislature can claim "if the money is wasted, the accountability scores will not improve, and sanctions will (ultimately) take care of people who are not producing". A number of members of the state legislature voting for KERA have been targeted by opposition groups for defeat. However, no member of the state legislature targeted for defeat by anti-KERA forces has (yet) been defeated where the KERA vote has been seen as a key component of that defeat.

The KIRIS accountability system created a visible means of public accountability for the school system. Like the football coach who finds his won-lost record publicized, the accountability scores for all schools in the Commonwealth are published in the two newspapers with statewide circulation and many local papers. Such an accountability system allows members of the state legislature and other citizens to know if schools in their community earned rewards or face sanctions. The rewards (and sanctions) given to schools are highly symbolic achievements that may reflect positively (or negatively) on an entire community. Thus, when the state grants awards to a particular school, they are not only giving employees a small cash bonus, they are giving a "pat on the back" to an entire community.

Negative Effects of the Assessment

1. The KIRIS assessment has limited use for assessing the educational progress of individual children. Parents (typically) want to know how **their** child is doing in school in relation to the child's own capabilities and sometimes in relation to peer groups in other communities or in other states. KIRIS is a school based accountability measure, looking at grade-on-grade changes in a school and does not track individual students over time. The current reliabilities (on the KIRIS assessment) "are not sufficiently high to make student level decisions without additional information" (Kentucky Department of Education 1994). It is possible that the KDE is moving to address the issue of inadequate student level reliability in future years of the KIRIS assessment.

BEST COPY AVAILABLE

The RFP issued in October 1995 will make it easier to create an assessment with improved student level reliabilities.

2. The corruptibility of high stakes assessment. Madaus (1988) notes that high stakes test scores may become the most important goal of education. Haerte (1994) pointed out an important component of education is lost when teachers and students work for higher scores on the KIRIS accountability index instead of the intellectual attainments an increasing accountability score is supposed to represent. With any recurrent high-stakes assessment, a tradition of past examinations develops, and, over time, examiners become reluctant to make significant changes from year to year because then teachers will not know what to teach. If that happens, the domain of assessment tasks will grow too narrow. Assessment scores will rise, but instruction will become stereotyped. As a consequence, students who have learned to do well on the particular kinds of items included in the assessment may do poorly on equally valid items that are assessing the same skills in a slightly different way.

3. The heavy concentration on writing and group work may detract from the development of other skills, especially basic skills. Resnick and Resnick (1991) have pointed out "you get what you assess and don't get what you don't assess". At the present time, the KIRIS assessment is oriented almost exclusively toward writing. The increase in KIRIS scores observed in the first three years of the program may reflect the heavy weight given to writing. Other skills are assessed in the KIRIS, but subsumed within writing. Alternative assessments of any type (authentic or multiple choice) stressing non-writing skills may show declines.

As mentioned earlier, the high stakes component of the KIRIS assessment is supposed to alter the way teachers teach. By stressing "higher order" cognitive skills (written responses to questions) it was thought that teachers would no longer "drill and kill" student to memorize bits and pieces of fragmented information. Examples of such bits and pieces of information would be memorization of multiplication tables and spelling. It was thought that "drill and kill" instructional strategies not only turn students off to learning, but focuses learning on a narrow set of skills that do not easily generalize to acquiring other information. Assessment like CTBS-IV and KEST may create in the minds of teachers the importance of bits and pieces of fragmented information in the curriculum. The reformers reasoned that teachers could not ignore the authentic assessment if stakes were associated with part of the system. However, the creation of the new assessment system presented at least two challenges to teachers, what to teach and when to teach it.

Teachers have a limited (and fixed) amount of time for instruction over an academic year. In that limited and fixed amount of time, teachers have to decide how to allocate their time to achieve instructional aims. Teachers in Kentucky, especially those teaching in the accountability years, need to allocate some of their instructional effort in teaching students to respond to essay questions and preparing students to develop portfolios. Teachers can turn any of the higher order cognitive skills assessed by KIRIS into a mechanical process by repeatedly teaching students the method of responding to essay questions. If teachers spend weeks (or months) training students to answer essay questions, it is possible that such students will not have learned enough information to master the content of a given subject. Such students may be able to master the *process* of doing well on the accountability assessment (KIRIS) and some of the higher-order thinking skills stressed by KIRIS, but perform poorly on skills not stressed by KIRIS. The KIRIS assessment does not stress some basic skills, like mathematical calculation.

Some evidence for a drop in basic skills comes from CTBS-IV test results from the largest school district in the Commonwealth of Kentucky. In a pre-test, post-test, quasi-experimental design, the pre-test was the average percentile rank of a grade of students in the Jefferson County Public School district before the KERA was passed (Campbell 1969, Cook and Campbell 1979). The post-test was the average percentile rank of a grade of students in Jefferson County Public District after the passage of the KERA. All components tests or the CTBS-IV showed decreases

BEST COPY AVAILABLE

FIGURE 1

Sixth grade math computation CTBS-IV

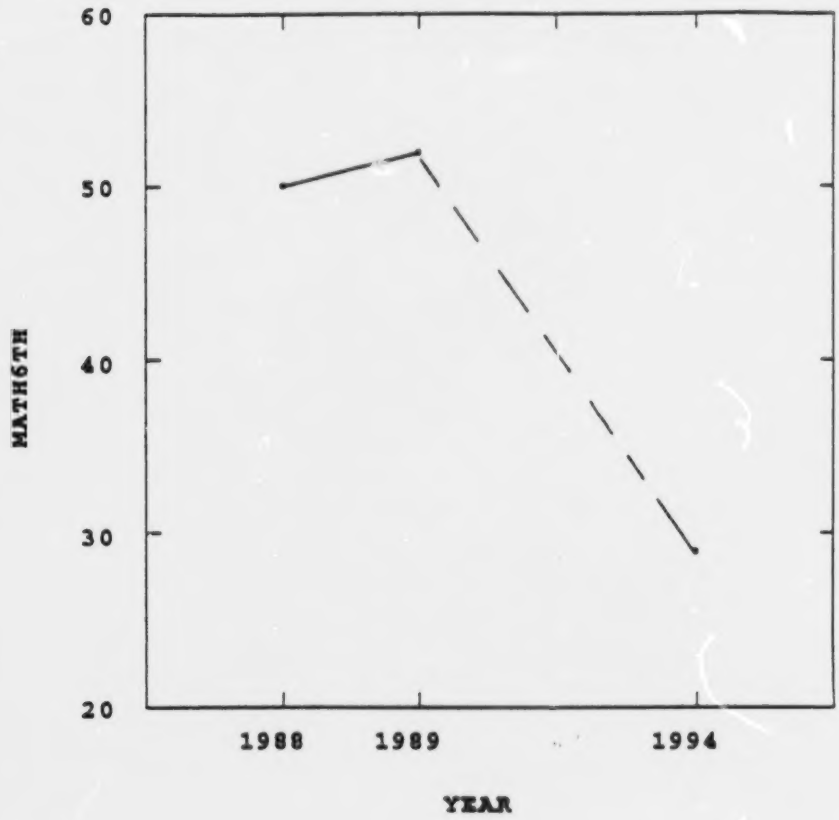
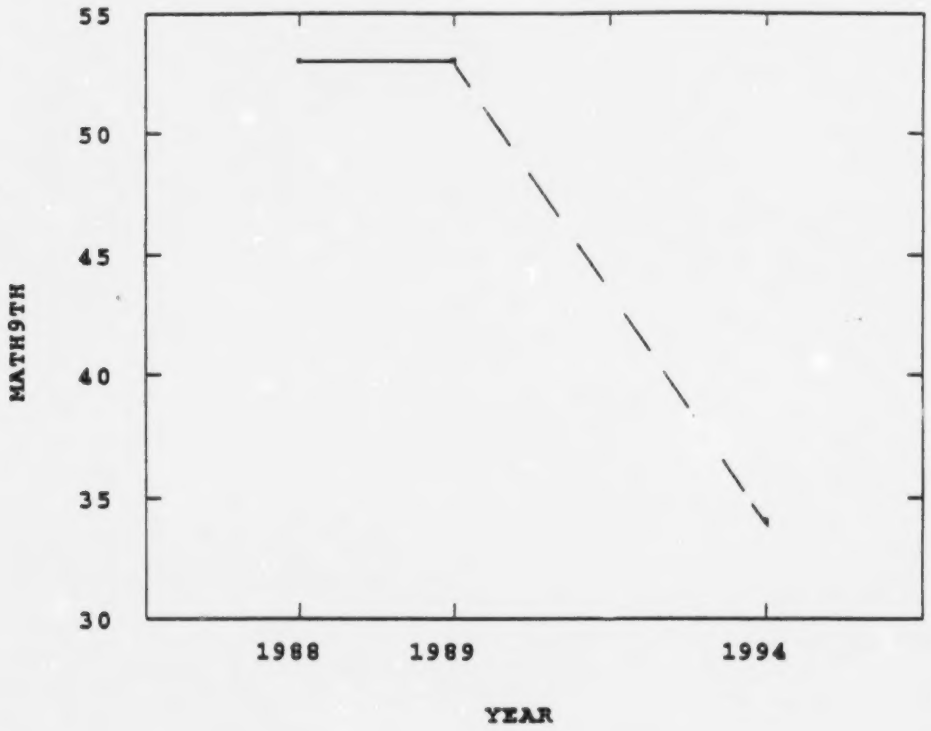


FIGURE 2

Ninth grade math computation CTBS-IV



in average percentile ranks after the passage of KERA, when compared to average percentile ranks before the passage of KERA. Decreases ranged from 8 percentile points to over 20 percentile points. The two largest drops, math computation scores for the 6th and 9th grades are presented in Figures 1 and 2, respectively. Again, the KDE may be deal with this issue when KIRIS is revised to meet the requirements of the October, 1995 RFP.

4. The public Does Not Yet Trust the Assessment. It is typical that stakeholders question the validity of new methods to demonstrate accountability. Initial skepticism is an appropriate response because it allows stakeholders to pose questions, seek rationales, and buy time until the new system provides data addressing legitimate concerns. This questioning demonstrates that stakeholders are taking an innovation seriously. Therefore, it is not surprising to find that stakeholders have serious concerns about KIRIS.

A recent study conducted by Wilkerson and Associates, Inc. (1994) for the Kentucky Institute for Educational Research found that principals, coordinators/supervisors, teachers, school council parents, public school parents, and the general public **all** ranked student performance on the KIRIS as the measure least likely to provide a reliable indicator of student learning. These diverse constituencies had most confidence in the percentage of students who finished high school. A study of the Kentucky state legislature found that 44 percent of the responding legislators said the most common complaint mentioned by the public was that the KIRIS was an inaccurate measure of students' abilities (Horizon Research International, 1994).

5. Overconcentration of assessment based on writing ability. The KIRIS assessment currently is heavily oriented toward evaluations based on writing. Students who have content knowledge of the discipline but lack adequate writing skills are precluded from doing well on the assessment. Some examples of alternative kinds of assessments that could lessen the impact of writing on the assessment include oral communication, involving the giving of a speech; creating a presentation that simultaneously involves written and visual information, typically called multimedia and usually done on a computer; and performing and fine arts. We note too that multiple-choice items offer students who lack adequate writing ability an alternative way to express the knowledge they have. Multiple-choice tests require good reading skills, however.

6. Portfolios will increase teacher workloads, producing increased stress. Vermont implemented a low stakes statewide portfolio assessment program during the 1991-92 school year. Teachers involved in the Vermont portfolio assessment reported portfolios as a worthwhile burden (Koretz, Stecher, Klein, & McCaffrey, 1994). However, these same teachers reported that portfolios caused considerable stress. Koretz, Stecher, Klein, and McCaffrey (1994) report

> The pressures experienced by educators went beyond time demands. For example, more than half reported difficulty finding appropriate tasks. Educators also reported feeling stress because of their uncertainty about appropriate uses of portfolio scores; the rapid implementation of the program; and inadequate, tardy, and inconsistent information from the state.

In focus groups held in Kentucky, we found that teachers thought that portfolios were beneficial to instruction. However, Kentucky teachers reported the same concerns as their Vermont counterparts. Increased stress was reported by nearly every teacher attending the focus groups. Additionally, a study of one Western Kentucky school district found that teacher stress is extremely high, approaching debilitating levels for many (Hughes & Craig, 1994).

Portfolios increase teacher workloads. This increase in teacher workload occurred in both low stakes assessment programs (Vermont) and a middle stakes assessment program (Kentucky).

BEST COPY AVAILABLE

It is *possible* that Kentucky teachers experience greater job stress than Vermont teachers due to the high stakes Kentucky assessment. However, we have no data on this question.

More General Problem: The Public Does Not Yet Trust Reforms Associated with Systemic Initiatives

From national surveys, we can conclude that the public does not object to replacing multiple-choice tests with more authentic assessments (Johnson and Immerwahr, 1994). Additionally, Johnson and Immerwahr (1994, P. 19) state

> Previous research by Public Agenda has suggested that large numbers of Americans, like leaders, question the usefulness of multiple-choice exams and favor alternatives such as essay tests, portfolios and demonstration projects when they are used in conjunction with grades. In this study, 54% of respondents say replacing multiple-choice tests with essay tests would improve academic performance-an endorsement, but one that falls significantly short of people's support for removing disruptive students (73%) or making correct English a requirement for graduation (88%).
>
> The problem that education reformers face in their drive to replace multiple-choice tests with more authentic forms of assessment is not that people object to the idea. The problem is that this particular recommendation seems somewhat tangential to people's chief concerns about the schools. It is as if people are saying, "Well, that's all well and good, but what about the guns, the drugs, the truancy, and the students who can't add, spell, or find France on a map?

Systemic reformers, including leaders of the Kentucky reform movement , seem to be at odds with the public's perception of what needs to be done in the schools. The public has three major concerns: order, discipline, and teaching the basics (Johnson and Immerwahr, 1994; Gallop Public Opinion Poll, 1994). Systemic reformers stress higher order cognitive skills, authentic assessment, enriched curriculum, and improved professional development of educators (Cohen, 1995). It seems that systemic reformers are not addressing the public's major educational concerns. According to Yankelovich (1995) educational leaders must engage the public's preoccupation with order, discipline and teaching the basics. Even after addressing the public's major concerns, leaders will need a mount an awareness campaign to increase support for systemic reform initiatives. Such a campaign can take a few years before public opinion surveys show increasing public support for the actions of educational leaders. In 1996, it is all to easy for politicians or other leaders to cripple a systemic reform initiative because such initiatives currently have (at best) very limited public support.

Conclusion

The KERA is one educational reform that has impacted instruction in the classroom for many students in Kentucky. The rewards and sanctions component of the assessment system made it very difficult for educators (teachers, principals, and superintendents) to ignore the reform act.

The act was designed to shock the system and force change. In many respects the system was shocked and change forced upon educators, perhaps due to the sanctions of the KIRIS assessment. However, shocking the system was a short-term strategy to move the system in a non-incremental manner. The problem in Kentucky is now to move from a new system that promised non-incremental educational change to a system that can fine tune itself without external shocks. A committee of the state legislature came within one vote of overhauling the KIRIS

assessment system in July 1995. It is not clear that the current system, sustained on rewards and sanctions, will last until 2012, the supposed end date of the program.

Acknowledgments

We wish to thank the Kentucky Institute for Education Research, especially Roger Pankratz and Nila Weddle, for all of the help they have given to the research team over the last seven months. Without their help, this report would have suffered severe shortcomings. With their help, we have been in a better position to avoid factual and interpretive errors.

We also want to thank the hundreds of other people throughout the Commonwealth of Kentucky who played a role in this study. Unfortunately, only a few of these people can be mentioned in a paper like this one.

During the qualitative data-gathering role, Robert Rodosky (Louisville), Duane Miller (Owensboro), Evonne Slusher (Bell County), and Joel Brown (Bowling Green) put together an impressive group of participants for the focus groups under a very tight timetable.

Many thanks to the Kentucky Association of Assessment Coordinators (KAAC) for allowing Fenster to "invite himself" to the May 20, 1994, meeting of the group. Because of that meeting, we added to the basic methodology of the study and decided to send surveys directly to stakeholder groups. If Fenster had not been able to attend that meeting, the idea for the survey would not have materialized. The surveys improved the quantity and quality of evidence presented in this report.

A special note of thanks to the 113 DACs and 70 superintendents who took the time from their busy schedules to answer an intensive questionnaire about their experiences with KERA and KIRIS. Without the time and effort of these people, the study would have been significantly weaker.

We thank Edward Reidy of KDE and Richard Hill and Amy Sosman of ASME for taking the time from their busy schedules to provide documents and to repeatedly answer our telephone questions on the KIRIS assessment.

We also recognize the long and hard work put into the KIRIS assessment system by ASME and KDE. Performance assessments are not yet "commonly" used. The problems with these new kinds of assessments have not yet been worked out technically nor operationally. It would have been easy for ASME and KDE to go slowly when implementing a new performance assessment system. ASME and KDE took the tougher road, bypassed the transitional testing period, and implemented the legislatively mandated performance-based system immediately.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Campbell, D. T. (1969). Reforms as Experiments. American Psychologist. 24:409-429.

Cohen, D. K. (995). What is the System in Systemic Reform? Educational Researcher. 24, No. 9, pp. 11-17, 31.

Cook, T. D., & Campbell, D. T. (1979). Quasi-Experimentation: Design and Analysis for Field Settings. Chicago, Il: Rand McNally.

Coe, P., Leopold, G., Simon, K., & Williams, J. (1994). Perceptions of school changes: Interviews with Kentucky students. A Report Submitted to the Kentucky Caucus of the Appalachia Educational Laboratory Board of Directors. Charleston, WV: AEL.

Gallop Public Opinion Poll. (1994). Gallop Survey of the Public's Attitudes Toward Public Schools. National Phone Survey of 1, 326 adults.

Guskey, T. R. (1994). High stakes performance assessment: Perspectives on Kentucky's educational reform. Thousand Oaks, CA; Corwin Press.

Haertel, E. H. (1994). Theoretical and Practical Implications. In T. R. Gusky (Ed.), High stakes performance assessment: Perspectives on Kentucky's educational reform. pp. 65-75. Thousand Oaks, CA; Corwin Press.

Horizon Research International. (1994). A survey of legislators on Kentucky instructional results information system (KIRIS). Legislative Research Commission, Office of Education Accountability. Frankfort KY: Author

Hughes, K,R., & Craig, J.R. (1994, November). Using performance assessment achievement data to evaluate a primary instructional program. Paper presented at the annual meeting of the American Evaluation Association, Boston, MA.

Johnson, J. and Immerwahr, J. (1994). First Things First What Americans Expect from the Public Schools. New York, NY: Public Agenda.

Kentucky Department of Education. (1994). Kentucky instructional results information system: 1992-93 technical report. Frankfort, KY: Author.

Kentucky Department of Education. (1993). Kentucky instructional results information system: 1991-92 technical report. Frankfort, KY: Author.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, S. (1994). The Vermont portfolio assessment program: findings and implications. Educational Measurement: Issues and Practices, 13(4), pp. 5-16.

Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. Educational Researcher, 23(9), pp. 4-14.

Madaus, G. F. (1988). The Influence of Testing on the Curriculum. In L. N. Tanner (ed.), Critical Issues in Curriculum (Eighty-seventh yearbook of the National Society for the Study of Education, pt. 1. pp. 83-121). Chicago: University of Chicago Press.

Resnick, L.B. and Resnick, D. P. (1991). Assessing the Thinking Curriculum: New Tools for Educational Reform. in B. G. Gifford and M. C. Conner (Eds.) Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction. (pp. 37-75). Boston: Kluwer Academic Publishers.

Rose v. Council for Better Education, Inc. KY 88-SC-804-TG (September 28, 1989).

Wilkerson & Associates, Ltd. (1994). Statewide education reform survey. Louisville, KY: Author.

Yankelovich, D. (1995). The Crisis in Education: Perspectives from the Public. Oral Presentation given at the Educational Policy Institute, Oakland University, September 7-8.